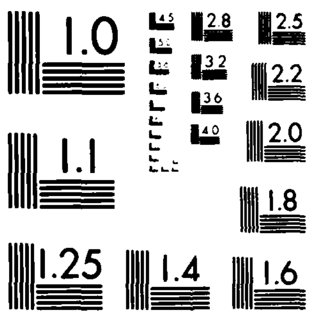
AD-A130 390 COMPUTERIZED AUDIO PROCESSOR(U) QUEENS COLL FLUSHING NY 1/1
DEPT OF COMPUTER SCIENCES M R WEISS ET AL. MAY 83
RADC-TR-83-109 F30602-80-C-0197

UNCLASSIFIED F/G 20/1 NL

END
DATE
FILMED
8 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

ADA130390

RADC-TR-83-109
Final Technical Report
May 1983

COMPUTERIZED AUDIO PROCESSOR

Queens College

Mark R. Weiss and Ernest Aschkenasy

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441

83 07 13 001

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-83-109 has been reviewed and is approved for publication.

APPROVED:

JOHN V. FERRANTE, Captain, USAF
Project Engineer

APPROVED:

THADEUS J. DOMURAT
Acting Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER:

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-83-109	2. GOVT ACCESSION NO. AD-A130 390	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPUTERIZED AUDIO PROCESSOR		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report Jul 80 - Aug 82
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) Mark R. Weiss Ernest Aschkenasy		8. CONTRACT OR GRANT NUMBER(s) F30602-80-C-0197
9. PERFORMING ORGANIZATION NAME AND ADDRESS Queens College/Dept of Computer Sciences Flushing NY 11367		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 64750F 11740043
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss AFB NY 13441		12. REPORT DATE May 1983
		13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: John V. Ferrante, Captain, USAF (IRAA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Enhancement
Narrowband Noise Suppression
Impulse Noise Suppression
Wideband Noise Suppression
Interference/Noise Reduction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Computerized Audio Processor is a computer synthesized electronic filter that removes interference from received or recorded speech signals. The CAP automatically detects and attenuates impulse sounds and tones. It also attenuates wideband random noise. All operations of the CAP are fully automatic. Input signals are processed in real time with a maximum lag of 340 msec.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Contents

1.0 Introduction 1
1.1 General Description of the Computerized Audio Processor 1
1.2 Background 2
1.3 Performance Specifications 4

2.0 The Control Panel of the Computerized Audio Processor ... 5
2.1 Input Control 5
2.2 Process Control 7
2.2.1 Impulse Attenuation 7
2.2.2 Tone Attenuation or Extraction 7
2.2.3 Wideband Noise Addition 7
2.3 Output Control 8

3.0 Operation of the Computerized Audio Processor 9
3.1 Turn-On Procedure 9
3.2 Setting and Controlling the Level of the Input Signal ... 9
3.3 Attenuation of Impulse Noise 10
3.4 Attenuation of Tones 10
3.5 Attenuation of Wideband Random Noise 11
3.6 Control of the Output Signal 13

4.0 Design Modifications 15
4.1 Adjustable Analysis Period 15
4.1.1 General Considerations 16
4.1.2 Implementation of Process Period Selectability 19
4.2 Modifications of the INTEL Process 21
4.2.1 Generation of the Cepstrum Threshold Function 24
4.2.2 Computation of the Average Noise Cepstrum 24
4.2.3 Cepstrum Threshold Scale Factors 26
4.3 Modification of the Output AGC 27
4.4 Implementation of a Permanent Program Memory 28

5.0 Recommendations for Improvement of the CAP 31
5.1 System Maintainabliity 31
5.2 System Utilization 32
5.3 Further Improvement of INTEL 32
5.4 Reduction of System Cost 33

Illustrations

Figure		Page
1	The Control Panel of the Computerized Audio Processor	6
2	Overlap Weighting and Processing of Input Signals	20
3	Operations in the INTEL Process	23

1.0 INTRODUCTION

1.1 General Description of the Computerized Audio Processor

The Computerized Audio Processor (CAP) is a computer synthesized electronic filter that removes interference from received or recorded speech signals. The CAP automatically detects and attenuates impulse sounds and tones (e.g., ignition noise, switching transients, whistles, chirps, hum, buzzes, FSK telegraphy, etc.). It also attenuates wideband random noise. All operations of the CAP are fully automatic. Input signals are processed in real time, with a maximum lag of 340 msec.

The CAP implements three proven signal processing techniques. One of these (IMP) virtually eliminates most loud impulse noises. A second technique (DSS) automatically detects tones and attenuates them by up to 46 dB. The third technique (INTEL) provides up to 18 dB attenuation of wideband random noise.

The CAP is very easy to use and requires very little attention by the user. Operation of the processor is initiated automatically within one second after the power is turned on. The user can select any combination of noise attenuation processes. He can also select the modes of operation that optimize these processes. Once the system has been set up as desired, the only other adjustments the user may need to make are of the input and output signal levels. The effectiveness of the selected attenuation processes can be checked when desired by use of a switch that permits the user to monitor either the

unprocessed input signal or the processed output signal. Input, output, and monitoring connections are conveniently available on the front panel of the CAP.

The CAP is composed of two units, both of which are contained in a frame that is 15 inches high, 19 inches wide, and 23 inches deep. One of these units, the system control unit (or SCU), contains the switches, potentiometers, and associated circuits that are used to control the operation of the CAP. It also contains an input signal conditioner that converts the analog input signal to digital form for input to the second unit, a Macro-Arithmetic Processor (or MAP), manufactured by CSPI. This device is a small, very powerful digital computer that performs the processing of the CAP input signals. The programs that implement the IMP, DSS, and INTEL processes are stored in an EPROM memory in the MAP. Digital output signals from the MAP are converted back to analog form by circuits in the system control unit.

The CAP requires 110 volt 60 Hz single phase power, and draws 12 amperes. Although designed for a laboratory environment, the CAP can operate reliably over an ambient temperature range of from 0 to 35 degrees centigrade.

1.2 Background

The CAP is the latest model in a series of speech enhancers, all of them developed under the support of the U. S. Air Force. Most of the earliest versions took the form of

computer programs that were run in very large digital computers. Typically, processing time took up to 40 times real time. The first practical, real-time implementation of the signal processing techniques was constructed in 1977. Known as an Advanced Development Model (ADM), it was used by the Air Force in a series of tests that were designed to determine the effectiveness and usefulness of this type of device. Based on the results of these tests, the ADM was modified to extend its performance and to simplify the controls that were available to the user. The modified instrument, now known as a Speech Enhancement Unit (SEU), was installed at an Air Force site where it has been used to process speech signals that were obtained under a wide range of practical conditions.

The experience gained during the construction and subsequent testing of the SEU have been incorporated into the design of the CAP, which differs from its predecessors in several important respects. Unlike the SEU, which required an external host computer to load the processor programs into it, the CAP is a self contained stand-alone unit. The signal processing range of the CAP, from 20 Hz to 3600 Hz, is 20 percent greater than that of the SEU. The effective amplitude dynamic range is 6 dB greater. Finally, the control panel of the CAP includes several additional controls that permit the user to achieve better optimization of its performance.

1.3 Performance Specifications

INPUT CHARACTERISTICS

Sensitivity:	2 volts rms maximum, 60 mv minimum for full dynamic range.
Input Impedance:	100,000 ohms
Input Connector:	BNC on front panel
Input Level Control:	Manual or AGC, selectable by switch on the control panel

OUTPUT CHARACTERISTICS

Signal Level:	Adjustable, 0 to 3 volts rms into 8 ohms
Frequency Range:	20 Hz to 3600 Hz
Output Impedance:	less than 2 ohms
Output Connector:	BNC
Output Level Control:	Manual with output AGC. Output AGC can be active or inactive (held) by use of selection switch on panel

PROCESSOR CHARACTERISTICS

Dynamic Range:	60 dB
Impulse Attenuation:	36 dB typical, 50 dB maximum
Tone Attenuation:	36 dB typical, 46 dB maximum
Wideband Noise Attenuation:	12 dB to 18 dB, depending on selection of cepstrum threshold (by use of switch on control panel)

2.0 THE CONTROL PANEL OF THE COMPUTERIZED AUDIO PROCESSOR

The CAP control panel is illustrated in figure 1. As can be seen, the controls are grouped according to function. In addition to controls, each group contains LEDs that function as system monitors and that indicate which control functions are active.

2.1 Input Control

The input signal is connected to the CAP via a BNC connector 1. The input impedance seen at this point is 100K ohms.

A 2-position toggle switch 2 provides the user with the ability to select the mode of control of the level of the input signal. With the switch in the AUTO position an automatic level control circuit maintains the signal at an optimum level at the input to the A/D converter. With the switch in MANUAL position, the user can adjust the level by use of a potentiometer 3.

A light emitting diode (LED) 4 turns on when the signal at the input of the A/D converter is within 20 percent of the maximum linear input to this circuit. For continuous input signals, this will occur when the input level control selector 2 is set to AUTO and the input peak level exceeds 3 volts. To maximize the dynamic range of the system when control is in the manual mode, level control 3 should be set such that the overload indicator 4 turns on occasionally.

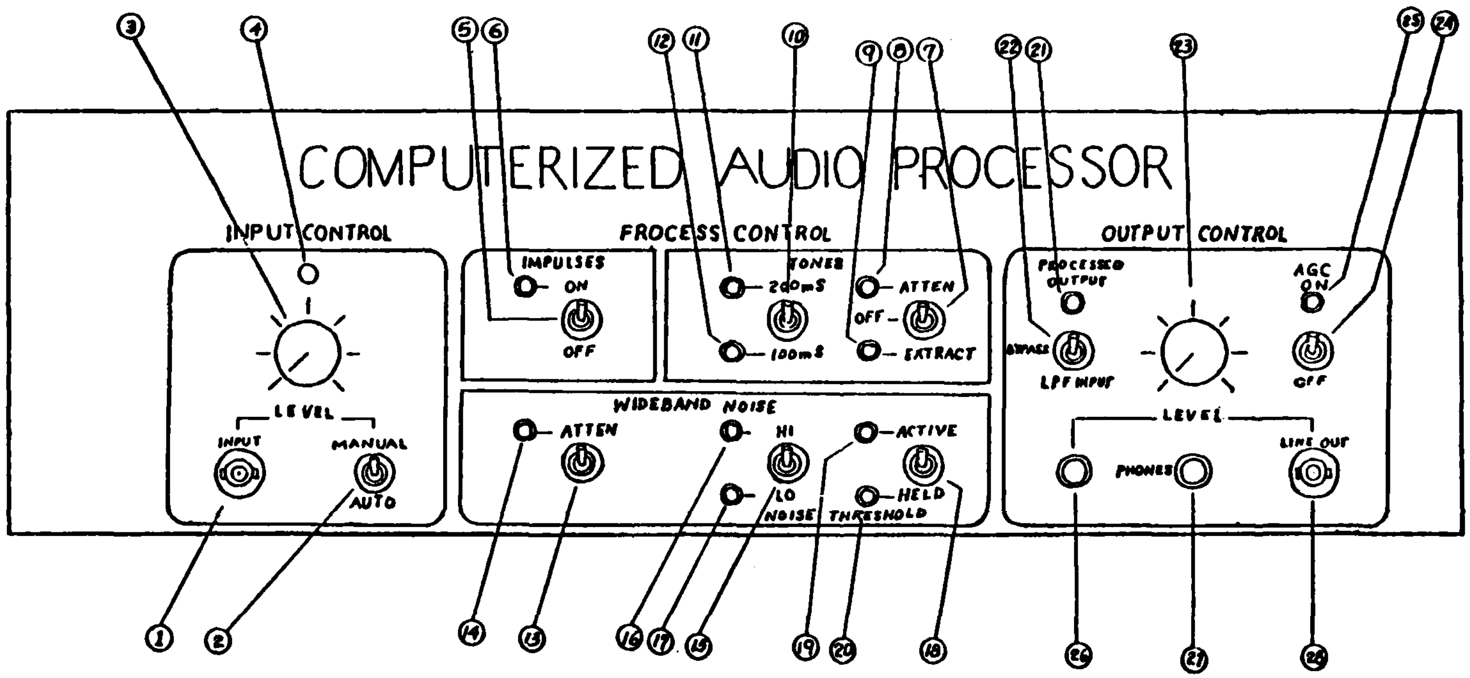


FIGURE 1 THE CONTROL PANEL OF THE COMPUTERIZED AUDIO PROCESSOR

2.2 Process Control

2.2.1 Impulse Attenuation

When process selection switch (5) is set to ON, the IMP process becomes active and LED indicator (6) will turn on.

2.2.2 Tone Attenuation or Extraction

The DSS process is controlled by a 3-position toggle switch (7). When this switch is in the middle position the DSS process is disabled. When the switch is set to ATTEN, indicator LED (8) will turn on, indicating that the DSS process is active and that it is detecting and attenuating tonal noises. When switch (7) is set to EXTRACT, indicator LED (9) will light, indicating that the process is detecting and extracting tones from the input signal.

For relatively stationary tones (those that change in frequency by less than 10 Hz/sec) the process duration switch (10) should be set to 200 ms (LED (11) will light). For less stationary tones the switch should be set to 100 ms.

2.2.3 Wideband Noise Attenuation

The INTEL process is made active when switch (13) is set to ATTEN. Indicator LED (14) will turn on. Switch (15) is used to select the cepstrum threshold level -- HI (indicated by LED (16)) or LO (indicated by LED (17)). The threshold is updated continuously when switch (18) is set to ACTIVE (LED (19) lights), or it is held constant when switch (18) is set to HELD.

2.3 Output Control

Switch 21 is a 3-position switch that selects which signal is to be delivered to the output connectors. When the switch is set to BYPASS, the unfiltered, unprocessed input signal is selected for output. When the switch is set to LPF INPUT, the low-pass filtered, but unprocessed signal is selected. In the PROCESSED position, the switch selects the signal that has passed through the MAP and indicator LED 22 will turn on.

The level of the output signal can be controlled by potentiometer 23 .

With switch 24 in the AGC ON position, indicator LED 25 will turn on, indicating that the level of the processed output signal is being maintained within a 16-dB range by the AGC routine in the MAP software. With switch 24 set to OFF, the AGC gain level is held constant at the value that existed at the moment the switch was set to OFF.

Output connectors 26 and 27 are standard phone jacks. BNC connector 28 provides a convenient line output for the Computerized Audio Processor.

3.0 OPERATION OF THE COMPUTERIZED AUDIO PROCESSOR

3.1 Turn-On Procedure

The CAP is turned on simply by pressing the Power On switch on the lower left side of the instrument. Within about one second the yellow light (labeled RESET) will begin to blink, indicating that the processor's programs have been loaded into the working memories of the MAP and that the CAP is ready for use. If startup of the CAP fails to occur automatically, the user can initiate the loading of the programs by pressing the RESET button.

3.2 Setting and Controlling the Level of the Input Signal

The level of the input signal can be controlled manually by use of the Input Signal Level potentiometer. It also can be controlled automatically by the input AGC circuits. The user can select either mode by setting the selection switch on the control panel to the appropriate position. Input AGC should be used when the level of the signal to be processed is in the range 60 mv rms to 2 v rms, since it will automatically adjust the signal level inside the CAP so as to make optimum use of the A/D conversion system. The user should select manual control when the level of the applied signal exceeds 2 v rms or when the signal-to-noise ratio at the input is greater than 20 dB. For either mode of control, the Input Signal Level potentiometer should be set such that the Overload indicator lights up occasionally when the input signal is at its peak level. However, when the input contains

frequent large impulses, use of the Overload indicator to guide the setting of the potentiometer can lead to the input level being made unnecessarily low when the manual mode of control is in use. The AGC mode of input level control is recommended for this type of signal.

3.3 Attenuation of Impulse Noises

The IMP process can be made active by use of the selector switch on the control panel. IMP will detect and remove from the input signal any impulses that are (1) larger than the peaks of speech sounds in the input signal, (2) no more than 8 msec in duration, and (3) spaced no closer than 15 msec from an adjacent impulse. IMP reduces to zero level those regions of an input signal in which impulses are detected. The gaps that this leaves in the signal are filled by overlapped and summed segments of the input signal adjacent to the deleted impulses.

3.4 Attenuation of Tones

The DSS process detects and identifies as tones those components of the input signal that have greater stability in frequency and amplitude than do components of speech. In general, the more stable that tones are in amplitude and frequency, the easier it is to distinguish between them and components of speech. For tonal noises such as constant pitch whistles, harmonics of power line hum, buzzes, etc., a 200-msec long segment of the input signal provides adequate ability to separate the components of such noises from those of speech.

However, the components of more dynamic tonal noises, such as heterodyne chirps and FSK telegraph signals, can be as variable as components of speech within a time window of 200 msec. For such noises it is necessary to shorten the time window to make such rapidly varying tones sufficiently stable within the time window and thereby make it possible for the DSS process to detect them.

The user is provided with the ability to set the DSS time window to either 200 msec (for relatively stable tones) or to 100 msec (for rapidly changing tones). The longer window usually results in a higher quality output signal and should be used whenever possible.

The primary use of the DSS process is to detect and attenuate tones in the input signal. On occasion, however, the tones may contain useful information. In such a situation it would be desirable to attenuate speech or other sounds that may be present and to extract the tones. A 3-position switch on the control panel provides the user with the ability to disable the DSS process, or enable it and either attenuate or extract any tones that are present in the input signal.

3.5 Attenuation of Wideband Random Noise

The INTEL process automatically reduces the level of random noise that may be present in the input signal. To perform this operation the processor continuously generates and updates the "cepstrum threshold". The cepstrum threshold is a reference pattern that contains in it a representation of the components of

random noise in the incoming signal. This pattern is generated at all times. When INTEL is made active, the process subtracts a scaled version of the cepstrum threshold from an equivalent pattern that represents the components of random noise plus any speech that may be present in the input signal. The result of subtracting these functions is a modified pattern in which the contribution of noise is reduced and that of speech is correspondingly increased. This modified pattern is used by INTEL to generate an output signal in which the noise level is much lower than it is in the input signal.

The user is provided with two options for control of the INTEL process. The first of these permits him to set the scaling of the cepstrum threshold to either a high level or a low level. A high threshold level results in maximum attenuation of random noise (up to 18 dB) and generally should be used when the input signal- to-noise ratio is very low (e.g., below 0 dB). The low threshold level results in moderate attenuation of random noise (up to 12 dB) and usually is appropriate for use when the input noise level is low to moderate. However, the user should compare the processed outputs for each setting and select the one that yields the highest quality.

The second control option permits the user to halt the updating of the cepstrum threshold pattern. This option should be exercised when the input signal has frequent dropouts that are longer than 0.5 seconds in duration. If the threshold pattern continued to be updated it would reflect the effect of a zero noise level during such dropouts. Consequently, when the signal

was restored the threshold pattern would be poorly matched to the noise distribution. This would result in an output signal noise burst that would gradually diminish as the threshold pattern was updated to reflect the presence of noise in the input signal. By exercising this option, the user "holds" the threshold pattern constant during a dropout and thereby keeps it matched to the noise in the input signal at the time the signal is resumed.

3.6 Control of the Output Signal

The output of the CAP is made available at a BNC connector and at two standard phone jack connectors. The user can select as the signal that is delivered to these connectors either (1) the original input signal (by putting the selector switch in the BYPASS position), or (2) the input signal after it has passed through the input and output anti-aliasing filters but without being passed through the MAP (LPF INPUT position of the selector switch), or (3) the signal that has passed through the MAP (PROCESSED OUTPUT position of the output selector switch). The signal level at the output connectors is sufficient to drive 2 watts into an 8-ohm load (e.g., headphones or a small loudspeaker).

In addition to a manual control for setting the level of the output signal the CAP provides automatic gain control for signals that pass through the MAP. This feature is made active by putting the AGC switch in the ON position. When the AGC switch is put in the OFF position, the output gain will be held at the value in use at the time the AGC was turned off. This

feature permits the user to establish a satisfactory AGC level and then to "freeze" it. This procedure should be used if the input signal contains frequent dropouts, since it will prevent the occurrence of bursting of the output signal level at the time the input signal is restored.

4.0 DESIGN MODIFICATIONS

The CAP is the latest and most powerful version of a speech signal enhancer that was developed for the U.S. Air Force under a series of contracts. The design of the CAP is based on that of the Speech Enhancer Unit or SEU, its immediate predecessor. To provide novel operating features that were required in the CAP, five major changes were made in the design of the SEU. These were (1) to make the DSS analysis period adjustable; (2) to make the level of the INTEL cepstrum threshold function adjustable; (3) to permit holding the cepstrum threshold function constant when necessary; (4) to permit holding the output AGC level constant when necessary; (5) to provide automatic high-speed loading of the system programs into the MAP from a permanent program memory.

The implementation of these features, the signal processing operations that were affected, and the performance improvements that were achieved are discussed in this section of the report.

4.1 Adjustable Analysis Period

Digital Spectrum Shaping (DSS) is the name given to that process in the CAP that removes sustained tonal noises from the input signal. Tones are considered to be sustained if they are present for at least 50 msec. These components coexist with those of speech in the time domain, but generally do not do so in the frequency domain. Consequently, the first step in the DSS

process is to compute the frequency transforms of successive segments of the input signal. This is accomplished in the MAP by use of a fast Fourier transform (FFT) algorithm. Each of the resulting amplitude spectrums of the input signal is examined by two algorithms that are designed to distinguish between components of tones and those of speech. Tones detected in an amplitude spectrum are suppressed in the corresponding complex spectrum. The complex spectrum of the signal is then converted back to the time domain by use of an inverse FFT. The regenerated segments of the input signal are recombined to form a continuous signal in which tonal noises will be found to be greatly attenuated.

4.1.1 Implementation of Process Period Selectability

To be effective, DSS must satisfy several requirements. It must be able to detect components of tonal noises. It must be able to remove a maximum of tonal energy from the complex spectrum while removing a minimum of speech energy. And it must be able to regenerate a speech sound that is maximally free of discontinuities and distortion. The first two of these requirements can be satisfied only to the degree to which speech components are separable from tonal components in the spectrum. The third requirement is affected by this same condition and also by the degree to which discontinuities in the output signal can be masked.

Any process for separating signals from noise does so by exploiting differences in their characteristics. Tonal noises

differ from speech sounds primarily in their greater stability in amplitude and frequency. Voiced speech sounds are quasi-stable over short periods, typically 25 msec or less. Tones that are separable from speech must be stable for significantly greater lengths of time. Fortunately, this requirement is met by most of the tonal noises that are encountered in the transmission, reception, or reproduction of speech sounds.

For maximum separation, the components of tones and of speech should occupy distinctly different regions of the spectrum. Tone components are represented by spectrum peaks that are centered at locations that correspond to the tone frequencies. Speech components, on the other hand, are distributed throughout the spectrum. Consequently, to minimize the overlap of speech and tones in the spectrum the energy in each tone should be concentrated into as narrow a spectrum range as possible. This can be accomplished in part by applying a suitable amplitude weighting function to the signal within each analysis window before computing its Fourier transform, and in part by selecting an optimum duration for the analysis window.

Amplitude weighting of a stable sinusoidal signal helps to concentrate into the primary spectrum peak the energy that otherwise would be distributed in the side lobes associated with that peak. This results in the reduction of the level of the sidelobes and, in roughly inverse proportion, the widening of the primary peak. Weighting the amplitude of the input signal complicates the generation of an unweighted output signal, since the inverse Fourier transform of the spectrum of a weighted

signal will be a signal that is weighted in the same manner. Consequently, it is desirable to use a weighting function that provides adequate suppression of sidelobes while making it easy to generate an unweighted output signal. These requirements are satisfied by the triangular weighting function that is used in the CAP. Using this function, the sidelobes immediately adjacent to the primary peak are reduced to a level 28 dB below the peak. Succeeding sidelobe levels diminish at a rate of 12 dB per octave resolution half-bandwidth (i.e., one-half the reciprocal of the duration of the analysis window). The simplicity of the method for generating an unweighted output is illustrated in figure 2. Obviously, any weighting function for which the sum of the upper half of the function and the lower half is unity could be used. Several such functions are available. Of these the triangular (or Bartlet) function provides the best compromise between widening of the primary peak and suppression of the sidelobes.

The second factor that affects the separability of speech and tone components is the duration of the analysis window. The window can be made short enough that both tones and speech appear to be stable within it. As the window is widened the spectrum peaks due to speech components will widen due to changes in their frequencies during the duration of the window. At the same time, the spectrum peaks of tones that are still stable within the widened window will become narrower. For a given window duration, P, the maximum rate at which the frequency of a tone can change without causing severe widening or distortion of the peak and spreading of the sidelobes is given as

$$df = 4/P^2 \text{ Hz/second}$$

For the 200 msec window that was used in the SEU, tones whose frequencies changed at rates less than 100 Hz/sec satisfi.d the criterion given above. However, noises such as rapidly changing heterodyne whistles and FSK telegraph tones did not. A shorter window is needed to accommodate tonal noises that change that rapidly or are of such short duration.

Two window durations are available in the CAP, 200 msec and 100 msec. The shorter window permits effective separation of speech components from tones whose frequencies change at rates of between 100 Hz/sec and 400 Hz/sec.

4.1.2 Implementation of Process Period Selectability

The DSS process period is made selectable by altering the length of the buffers in which samples of the input signal are stored. To permit real time operation of the CAP with no loss of input data, the samples are stored in two buffers. Each buffer is loaded during the time that the contents of the alternate buffer are being processed, as illustrated in figure 2. For the nominal 200-msec processing period, each buffer contains 1024 samples, i.e., one-half the data in an analysis window. For the 100-msec processing period, the buffers are shortened to 512 samples.

Two of the component operations in DSS are adjusted to correspond to the length of the selected analysis period. One of these is the weighting of the analysis window. The length of the weighting function is adjusted to match the length of the

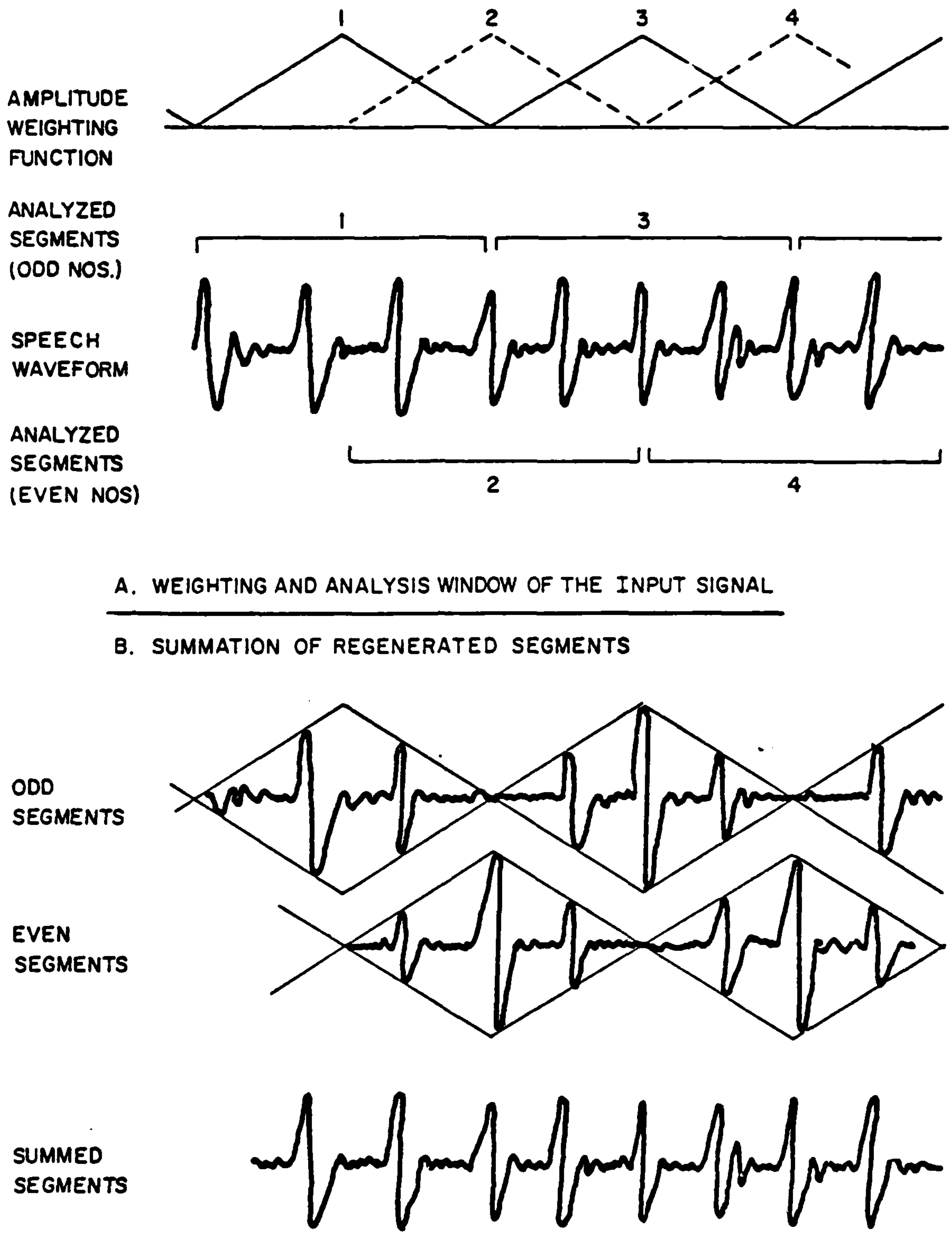


FIGURE 2 OVERLAP WEIGHTING AND PROCESSING OF INPUT SIGNALS

analysis window and the slope is adjusted so that the function begins with a value of zero, rises linearly to a value of unity at the center, and falls linearly to a value of 1/N at the end (N is the number of samples in the analysis window). The second operation is the detection and suppression of tonal components in the signal's spectrum. The algorithms that detect tonal components and the select zones in which these components will be attenuated are frequency based. That is, the various tests and procedures required for execution of the algorithms are performed within specified frequency bands and use specified frequency increments. Both of these are expressed in terms of numbers of samples. For a 200-msec analysis window the 5000 Hz wide spectrum is defined at 1024 uniformly spaced points, resulting in a frequency interval of 4.88 Hz per spectrum sample. The frequency interval is twice as great, or 9.76 Hz per sample, for a window length of 100 msec. It is necessary to maintain constant bandwidths and intervals in the DSS processes regardless of which analysis window is chosen. This is accomplished by making the number of spectrum samples required to define a frequency band or frequency increment for the 100 msec window half the number that is required for the 200 msec window.

4.2 Modifications of the INTEL Process.

The INTEL process is used in the CAP to attenuate additive wideband random noise that may accompany speech signals. Unlike impulses and steady tones, this type of noise will coexist with the speech signal continuously both in time and frequency.

Consequently, speech and additive wideband random noise cannot be separated from one another in either the time or the frequency domain. To achieve some degree of separation it is necessary to transform them to a new domain, that of the cepstrum.

In INTEL, cepstrum transformation is achieved by computing the amplitude spectrum of the square-root amplitude spectrum of input signals, as illustrated in figure 3. Although it is not strictly correct to call this new domain the cepstrum (which formally, is the power spectrum of the log amplitude spectrum of the input signal), we do so here for convenience. As in the true cepstrum, the transform represents the period content of the input signal, expressed in units of time, and described as quefrencies.

The effectiveness of the INTEL process lies in the way that it exploits differences in the cepstrum characteristics of noise and speech. Both types of signals yield cepstrum waveforms with the maximum amount of energy concentrated into the low quefrency region, below 0.6 msec. However, the percentage of total energy concentrated into this region is much greater for noise than it is for speech. Above 0.5 msec, noise energy falls continuously with increasing quefrency. Speech energy diminishes in a similar manner, but with significant local increases in energy at quefrencies that correspond to integral multiples of the pitch period.

Signals are processed in the cepstrum domain in such a way as to enhance the signal-to-noise ratio in the amplitude spectrums of incoming signals. This is accomplished in three

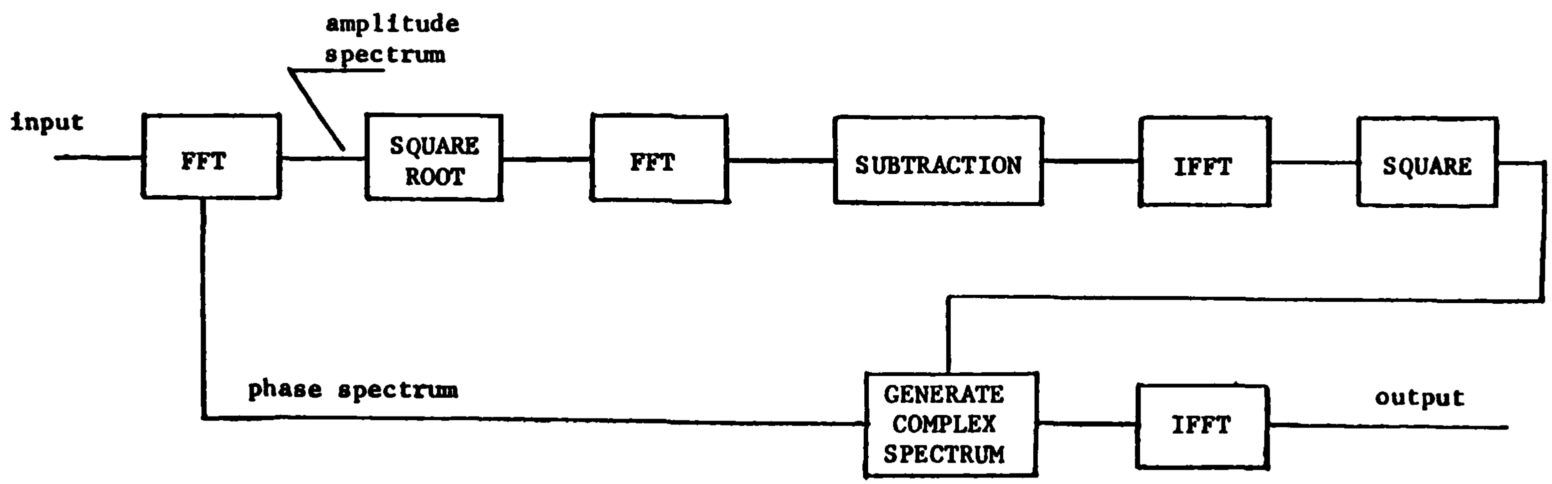


FIGURE 3 OPERATIONS IN THE INTEL PROCESS

steps. First, a scaled version of the average cepstrum of noise alone is computed. This function (the cepstrum threshold) is subtracted from the cepstrums of combined speech and noise. Then the inverse Fourier transform of the modified cepstrum is computed. Finally, the resulting modified square-root spectrum is squared to restore the correct relative amplitudes of the largest spectrum components.

4.2.1 Generation of the Cepstrum Threshhold Function

4.2.2 Computation of the Average Noise Cepstrum

Ideally, the cepstrum threshold function should reflect the current distribution of noise in the cepstrum. While it is not possible to achieve this goal in practice, it can be approached closely. This is accomplished by computing a lossy moving average of the distribution of noise in the cepstrum. The procedure that is used first determines if the input signal contains detectable voiced speech sounds. If it does not, then the cepstrum of whatever noise is present at the input is weighted by a factor W1, and added to the current average noise cepstrum, which is weighted by a factor W2. The resulting function becomes the new average noise cepstrum. The weighting factors are chosen such that the time-constant of the averaging process is about 0.5 second, a period that is short enough to allow the cepstrum threshold to follow moderately fast changes in noise distribution and long enough to permit the generation of a smooth average noise cepstrum when the noise distribution is stable.

For one class of input signals, the use of a lossy moving average can lead to the production of brief but severe degradations in the quality of the INTEL output. This will occur when silent intervals, or intervals when the signal level is very low occur in an otherwise continuous noisy input signal. Such intervals or gaps can occur when weak signals are detected by an FM receiver. The receiver output will contain a high level of noise during the time that the signal is strong enough to inhibit operation of the squelch circuit. However, when the signal becomes sufficiently weak (as through fading) the squelch will become operational and the receiver output signal will become very small. During such a gap the level of the average noise cepstrum will diminish. If a gap is one second or longer in duration, the level of this function will be so low that when the noisy signal reappears, the threshold must again be built up. For the first few tenths of a second the output of the INTEL process will be many times greater that it was just before the gap occurred. Consequently, a loud noise burst will appear in the INTEL output immediately after the signal reappears. This was extremely objectionable to users of the preceding version of this system.

The problem described above is eliminated in the CAP by providing the user with the ability to "hold" the average noise cepstrum constant when necessary. When the selection switch on the CAP control panel is set to the normal position (down), the average noise cepstrum is updated as described above. When the switch is set to the HOLD position (up), updating of the average

noise cepstrum ceases and the function that existed just before the switch was set is used until the switch is reset to the down position and normal updating of the average noise cepstrum is resumed.

4.2.3 Cepstrum Threshold Scale Factors

The cepstrum threshold function is a scaled version of the average noise cepstrum. Three scale factors are used in the generation of this function: one at zero quefrency, one for the cepstrum range between 0.1 and 0.5 msec, and the third one for the cepstrum above 0.5 msec. The optimum values for these factors have been determined empirically for a variety of noise distributions and S/N's. These are the values that yield the greatest reduction in noise level for the least distortion in the quality of the regenerated speech sounds or of the residual noise in the output signal. These values appear to be substantially independent of the distribution of noise in the spectrum. However, they do depend to some extent on the relative levels of speech signal and noise at the input. The lower the input S/N is the higher the optimum scale factors become. A single set of scale factors was provided in the earlier version of the CAP. These were determined for a S/N of about 3 dB. Two sets of scale factors are provided in the CAP, one of them for a S/N of 6 dB or greater, the other for S/N of 0 dB or less. Both sets of scale factors result in the same degree of enhancement and quality of the output signal for S/N in the range 0 dB to 6 dB. As described in Section 2.2.3, the selection of the desired set of

scale factors (i.e., cepstrum threshold level) is made by use of a switch on the control panel of the CAP.

4.3 Modification of the Output AGC

The level of the signal at the output of the CAP is controlled by an automatic gain control (AGC) program that attempts to maintain the peak signal amplitude within a specified range of values. Each buffer-length segment of the output signal is processed separately by the output AGC. First the samples of the output signal are compared to high and low amplitude limits. Then they are then scaled by an AGC gain factor that adjusts their magnitudes so as to keep the signal level at the output of the D/A converter within a desired voltage range. So long as the peak amplitude of the signal falls between the high and low amplitude limits, the AGC gain factor is held constant. If the peak level exceeds the high limit the gain factor is reduced by 6 dB. If it falls below the low limit the gain factor is increased by 6 dB.

For at least one condition -- when extended silent periods occur in the input -- it is desirable that the gain factor not vary with the level of the processed output. During such periods the signal level will continuously fail to exceed the low limit and so the gain factor will reach a very high value. Consequently, when the input signal level is restored, the initial output will be extremely large, i.e., a burst of signal. To prevent such an occurrence, the user of the CAP is provided with the ability to "freeze" the AGC gain factor. When the AGC

control switch on the CAP control panel is set to the AGC HOLD position, the gain factor is held constant at the last value that was in use before the switch was set. However, the level of the processed output signal samples will continue to be compared to the upper and lower amplitude limits and a "dummy" AGC gain factor will be established in the same manner as was used for the actual gain factor. When the AGC control switch is reset to the normal position, the value of the dummy AGC factor is transferred to the actual one. This procedure insures that when normal AGC operation is resumed the AGC gain factor will be correct for the level of the output signal present at that time.

4.4 Implementation of a Permanent Program Memory

The programs that constitute the CAP system software must be loaded into the MAP each time the system is to be used. In earlier implementations, the programs were stored permanently in either punched paper tape, magnetic tape, or magnetic disc. Consequently, in addition to a MAP, the system hardware included an input device (such as a paper tape reader, a tape drive, or a disc drive) and some means of controlling the input device, reading the programs from the permanent storage medium, and transferring them to the MAP. Usually a minicomputer was used for this purpose. Depending on the form of permanent storage used, the time taken to enter control commands into the minicomputer and then to load the programs into the MAP could take from a half minute to several minutes. Occasionally, read errors or equipment malfunctions made it necessary to repeat the

loading operations several times before a successful transfer of the programs was achieved.

In the CAP, the system programs are stored in an EPROM memory. The loading of the MAP is initiated automatically whenever power is applied to the MAP. Usually this will be when the main power switch on the MAP is pressed on. However, the system programs also will be automatically loaded into the MAP when line power is restored after a power failure has occurred. Loading of the MAP is initiated when a circuit on the I/O scroll detects a new application of line power. This circuit automatically generates a bus reset that clears all MAP memories and then it forces the starting address of a bootstrap loader program (that also is stored in the EPROMs) into the program counter of the CSPU in the MAP. The MAP then proceeds to load the bootstrap, which causes the main loader program to be loaded from the EPROMs. This program then loads the CAP system programs into the memory on bus 1 of the MAP. All loading operations are completed in less than 0.5 second. Thus, within 0.5 second after power is turned on the CAP is ready for use.

The permanent program memory resides on the I/O scroll. It consists of eight INTEL type 2716 ultraviolet erasable PROM integrated circuits. Each IC can store 2048 bytes of data. The 16K-byte capacity of the memory (8K half words) is substantially greater than the space required to store the CAP system software. The available unused storage capacity could be used to store test signal data, or system checking programs, or, as they are developed, new signal processing programs.

The use of EPROMs as the program storage element makes it very easy to modify the stored programs when desired. A 20-minute exposure to intense ultraviolet light is sufficient to completely erase the contents of the ICs. Following this, the modified programs are loaded into the MAP, usually from punched paper tape. A toggle switch on the I/O scroll provides the user with the ability to select either the EPROMs or an external device (e.g., a paper tape reader) as the source of the system programs. The erased EPROMs are then mounted, one pair at a time, into an EPROM burner which is connected by a ribbon cable to the I/O scroll. Successive 2K half-word segments of the system programs are loaded into different pairs of EPROMs. When the loading of a pair of EPROMs is completed the newly stored contents are read back and verified by comparing them with the corresponding segement of the program in the MAP.

5.0 RECOMMENDATIONS FOR IMPROVEMENT OF THE CAP

The CAP is by far the most effective version of its class of speech enhancement devices, and the easiest to use as well. However, there are several ways that it could be improved to make it more dependable, more effective, more usable, and possibly less expensive. These are discussed in this section of the report.

5.1 System Maintainability

While the electronic systems and circuits that make up the CAP are highly reliable, (the MTBF of the equipment is about 2800 hours) it is desirable to be able to verify conveniently and quickly that the system is operating correctly. When a failure does occur, down time can be minimized if the cause of the failure can be isolated quickly. The first of these objectives requires the ability to generate a set of standard input signals that can be used to test all of the operational features of the system. The system response to these signals could be verified manually or, if desired automatically, by comparisons with responses obtained when the system was known to be operating correctly.

Diagnostic testing of the system can be accomplished largely through the use of programs such as those that are used by CSP, Inc. to test the operation of the MAP. Similar programs can be written to test the operation of the system control unit and to isolate the causes of failures in that device.

To provide maximum convenience in testing the CAP, the necessary test signals and the diagnostic programs all can be stored in the permanent program memory on the I/O scroll. The memory can easily be extended, if necessary, by adding pairs of EPROMs. A switch can be added to the control panel that would permit the user to select between normal operation, performance testing, and diagnostic testing of the CAP.

5.2 System Utilization

At the present time, the CAP can be used to process only one input signal at a time in real time. The DSS and INTEL processes each take about half the available time. Together they consume about 98 msec out of the 102.4 msec period that is available when the nominally 200-msec analysis window is used. It is possible to modify the design of the system to permit two different signals to be processed at the same time, each one using only one of these two processes. Thus, two users could share the CAP when the signals they are monitoring contain either tones or wideband random noise. Since very little time is required by the IMP process, it could be made available to both users together with either DSS or INTEL.

5.3 Further Improvement of INTEL

The newly added ability to hold the cepstrum threshold function constant during abrupt signal dropouts proved to be highly effective in tests using both simulated and real signals. However, when the hold is released this function may not be

correct for the noise then present at the CAP input. This can occur because the average noise cepstrum is not updated during the period the cepstrum threshold function is held constant. The result can be a short moderate burst or muting of the output signal. The INTEL program can easily be changed to provide updating of the average noise cepstrum at all times. This will require a small expansion of the memory on bus 2 of the MAP to permit storage of both the updated average noise cepstrum and the one then in use.

5.4 Reduction of System Cost

The MAP is both the most costly and the largest component of the CAP. It is used because it can compute FFTs with 32-bit floating point precision and with sufficient speed. However, it is possible that a 16-bit fixed-point FFT computation would provide sufficient accuracy if appropriate block-scaling techniques were used. Fixed-point FFT devices generally are smaller and much less expensive than array processors such as the MAP. It is almost certain that a 16-bit range would be sufficient for IMP and DSS processing of virtually all the signals that are likely to be encountered in practical applications. However, it is not at all apparent that it would be adequate for INTEL processing of signals at S/N below 3 dB. The use of a high cepstrum threshold level for such signals can lead to greatly reduced levels in the regenerated amplitude spectrum of speech. Consequently, there could be significant levels of quantization noise in the output of a fixed-point "NTEL

process. The possibility of using this approach, the resulting degradations in speech quality, and the evaluation of ways to minimize them could be determined by simulating the process in a digital computer.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

DATE
ILME